

Un(e) apprenti(e) Data Scientist

- Niveau du diplôme préparé : Bac + 5 et plus
- Durée du contrat : 10 à 12 mois
- Emploi ouvert aux personnes en situation de handicap
- Poste localisé à Montpellier
- Poste à pourvoir en septembre 2024
- Domaines : *Numérique et digital, Big data statistiques, Systèmes d'information, Intelligence artificielle*

Description de l'employeur

La mission principale de l'Abes est d'aider les bibliothèques universitaires et de recherche françaises à décrire et rendre accessibles leurs ressources documentaires (livres, revues, articles, au format papier ou numérique).

Cette mission se traduit par la conception et la maintenance de grandes bases de données, décrivant des millions de livres, d'articles, de thèses, mais également des millions de personnes (auteurs, traducteurs, etc.), d'organismes, de concepts...

Ces grandes bases de données abritent des masses de métadonnées très structurées, de qualité, produites par des milliers de catalogueurs depuis vingt ans mais également par les éditeurs eux-mêmes. Ces métadonnées sont accessibles via les interfaces publiques sudoc.fr et theses.fr, mais également sur le web de données (*linked data*).

Face à l'explosion du nombre de publications scientifiques et aux enjeux de l'ouverture des données (qui rend nécessaire l'interopérabilité), l'Abes est à la recherche de nouveaux moyens pour automatiser certaines opérations et enrichir ses données. La science des données, et en particulier le **machine learning**, sont des outils aujourd'hui indispensables pour l'Abes.

Description du poste

En tant qu'apprenti-e data scientist, vous serez intégré-e au Labo de l'Abes. Cette structure est dédiée à l'innovation et s'est donnée pour priorité **l'exploration et l'expérimentation des solutions offertes par l'intelligence artificielle** au service des missions de l'Abes. Ce positionnement offre l'opportunité de s'approprier différentes technologies, pour contribuer à résoudre des problèmes variés, et ce en participant à l'intégralité d'un projet, du début à la fin, quelle que soit son envergure.

Au sein du labo, vous travaillerez avec l'équipe composée d'un informaticien senior et de deux data analysts seniors. En fonction des projets, vous serez aussi amené-e à collaborer avec d'autres équipes de l'Abes (notamment avec les informaticiens de la DSI) ou avec des prestataires extérieurs spécialisés en IA.

Les projets auxquels vous contribuerez s'inscriront dans la dynamique globale de l'Abes, ce qui est source de motivation et d'engagement collectif. L'activité de Recherche & Développement, incarnée par le labo et dédiée à l'IA, est une des priorités du nouveau projet d'établissement de l'Abes.

A ce stade, nous ne proposons pas de sujet ferme, mais nous veillons toujours à ce que les étudiants travaillent principalement sur un sujet précis et valorisant, même s'ils peuvent être sollicités sur d'autres tâches de l'équipe. Dans tous les cas, la mission portera sur des tâches relatives au **traitement du langage naturel**, en mobilisant **l'état de l'art de l'IA** le plus récent (*machine learning, deep learning, LLM, etc.*). Conformément à nos axes prioritaires, il s'agira **soit de travailler sur le texte intégral des thèses** (extraction d'information dans 100 000 fichiers PDF volumineux), **soit de travailler sur les métadonnées bibliographiques** (extraction, classification, etc.). Dans ce dernier cas, il est possible qu'une des tâches consiste à **intégrer dans une application Android l'appel à des LLM et autres modèles d'IA**. Toutes ces matières présentent des défis techniques et métier motivants.

Qui sommes-nous ?

L'Abes - Agence bibliographique de l'enseignement supérieur est un établissement public administratif (EPA) employant 90 agents - titulaires et contractuels - dans les filières bibliothèque, informatique et administrative.

En tant qu'agence nationale, l'Abes administre une gamme d'outils et services mutualisés au profit de l'ensemble des établissements documentaires de l'Enseignement Supérieur et de la Recherche :

- **Sudoc** : base de données bibliographiques pour faciliter l'accès aux collections des bibliothèques universitaires et de recherche
- **Thèses de doctorat** : gestion des métadonnées et de l'archivage des thèses numériques soutenues en France
- **Calames - archives et manuscrits de l'ESR** : description et accès aux collections d'archives et manuscrits de l'ESR
- **IdRef - identifiants et référentiels** : notices d'autorités produites par les bibliothèques de l'ESR
- **Bacon - Base de Connaissance Nationale et scienceplus.abes.fr** : métadonnées des ressources numériques à la disposition des fournisseurs d'outils de gestion de ressources électroniques et des établissements
- **Acquisitions de ressources documentaires numériques** : licences nationales, groupement de commandes permanent.

En savoir plus...

Gardons le contact !



Principales activités

- Analyse des données
- Analyse de l'état de l'art pour identifier les bonnes solutions
- Implémentation Python des solutions choisies, à adapter aux données et besoins Abes
- Conception et implémentation de métriques pour établir la pertinence des algorithmes
- Documentation

Descriptif du profil recherché

- Vous aimez explorer et exploiter des données textuelles et structurées.
- Vous souhaitez vous spécialiser dans le machine learning et le traitement du langage naturel.
- Vous faites preuve de curiosité et avez le goût de la recherche de solutions innovantes.
- Vous appréciez le travail en équipe.
- Vous avez une bonne maîtrise de Python (ou Java selon le projet) et une certaine maîtrise de SQL.
- Vous n'avez pas de difficulté à lire de la documentation technique en anglais.

Avantages sociaux

- Restauration : Tickets restaurants
- Accès aux prestations du CE, aux prestations sociales (tickets cinémas, chèques vacances, sorties, action sociale interministérielle...)
- Remboursement partiel des abonnements de transport domicile – travail
- Remboursement partiel de la cotisation complémentaire santé

Modalités de travail

- Semaine d'accueil et formations internes ciblées pour les activités du poste
- Accès à la formation permanente, à des congrès professionnels
- Organisation flexible des horaires de travail en accord avec le chef de service, dispositif de gestion des récupérations,
- Congés payés réglementaires et 5 jours ouvrables pour la préparation de l'examen
- Le télétravail pourra être accordé de manière ponctuelle

Informations complémentaires

- Accès parking fermé, local à vélos
- Proximité ligne de Tram

CV et lettre de motivation à envoyer à apprentissage-labo@abes.fr ou par courrier, adressé à :

*Monsieur le Directeur de l'Abes
Agence bibliographique de l'enseignement supérieur
227, Avenue du Professeur Jean-Louis Viala
CS 84308
34193 Montpellier Cedex 5*

Téléphone (standard) : 04.67.54.84.10

Contacts pour informations :

Yann Nicolas, responsable du labo (nicolas@abes.fr)

Sandra Beffaral, responsable du Service des ressources humaines (beffaral@abes.fr)

Les entretiens se déroulent à Montpellier ou en visioconférence